

# **AJUSTES DE FUNÇÕES DE DISTRIBUIÇÃO DE PROBABILIDADE AOS INSUMOS DE GERAÇÃO DE ENERGIA RENOVÁVEL**

**Jeane silva souza** (UFCG) - jeane.souza@ee.ufcg.edu.br

**Pollyanne De Oliveira Carvalho Malaquais** (GEPEN) - pollyannem7@gmail.com

**Benemar Alencar de Souza** (UFCG) - benemar@dee.ufcg.edu.br

**Washington Luiz Araújo Neves** (Instituição - a informar) - waneves@dee.ufcg.edu.br

## **Resumo:**

*Neste trabalho é apresentada uma metodologia para ajuste de distribuições de probabilidade à insumos de geração de fontes de energia renováveis (eólica e solar). Com base em análises estatísticas da função de distribuição de probabilidade e no teste de aderência Kolmogorov-Smirnov. A metodologia proposta é testada e os resultados são apresentados para quatro cidades da Paraíba (Brasil). Os resultados mostraram que, dentre as distribuições de probabilidade citadas na literatura, a distribuição de Valores Extremos Generalizados é muito flexível e pode modelar dez anos de dados de irradiância solar média global com valores-p superiores a 0,2 em todas as cidades analisadas. Para a velocidade do vento a distribuição Weibull é a que melhor representa a base de dados utilizada com valores-p acima de 0,07. A principal contribuição deste trabalho é apresentar que essa metodologia pode ser desenvolvida para orientar estudos que exijam dados de fontes de energia renováveis das regiões estudadas da Paraíba ou de outras localidades. Por exemplo, para realização de simulações generalizadas sobre fluxos de carga, dimensionamento e localização de sistemas. Ademais, pode ser replicada para diferentes bases de dados e outras variáveis.*

**Palavras-chave:** *Análises Estatísticas, Valores Extremos Generalizados, Weibull.*

**Área temática:** *Radiação Solar*

**Subárea temática:** *Outros Dados Meteorológicos aplicados à energias renováveis*

# AJUSTES DE FUNÇÕES DE DISTRIBUIÇÃO DE PROBABILIDADE AOS INSUMOS DE GERAÇÃO DE ENERGIA RENOVÁVEL

Jeane Silva de Souza – jeane.souza@ee.ufcg.edu.br

Pollyanne de Oliveira Malaquias – pollyanne.malaquias@ee.ufcg.edu.br

Benemar Alencar de Souza – benemar@dee.ufcg.edu.br

Washington Luiz Araújo Neves – waneves@dee.ufcg.edu.br

Universidade Federal de Campina Grande, COPELE, CGEE, Departamento de Engenharia Elétrica

**Resumo.** Neste trabalho é apresentada uma metodologia para ajuste de distribuições de probabilidade à insumos de geração de fontes de energia renováveis (eólica e solar). Com base em análises estatísticas da função de distribuição de probabilidade e no teste de aderência Kolmogorov–Smirnov. A metodologia proposta é testada e os resultados são apresentados para quatro cidades da Paraíba (Brasil). Os resultados mostraram que, dentre as distribuições de probabilidade citadas na literatura, a distribuição de Valores Extremos Generalizados é muito flexível e pode modelar dez anos de dados de irradiância solar média global com valores- $p$  superiores a 0,2 em todas as cidades analisadas. Para a velocidade do vento a distribuição Weibull é a que melhor representa a base de dados utilizada com valores- $p$  acima de 0,07. A principal contribuição deste trabalho é apresentar que essa metodologia pode ser desenvolvida para orientar estudos que exijam dados de fontes de energia renováveis das regiões estudadas da Paraíba ou de outras localidades. Por exemplo, para realização de simulações generalizadas sobre fluxos de carga, dimensionamento e localização de sistemas. Ademais, pode ser replicada para diferentes bases de dados e outras variáveis.

**Palavras-chave:** Análises Estatísticas, Valores Extremos Generalizados, Weibull.

## 1. INTRODUÇÃO

O Brasil, assim como outros países, incentivou o uso de novas fontes para a geração de eletricidade (solar, eólica, biomassa, entre outras). Esse crescente incentivo é impulsionado por diversos fatores, dentre eles tem-se a necessidade de maior confiabilidade, maior eficiência e geração de energia mais limpa. O conceito de geração distribuída (GD) surgiu como uma alternativa ao modelo tradicional de geração centralizada e tem papel fundamental na recentemente iniciada popularização das fontes de energia renovável no país. No entanto, o fornecimento de energia elétrica a partir de algumas fontes renováveis, como a energia solar fotovoltaica e a energia eólica, é estritamente dependente de fatores climáticos, por exemplo, irradiância solar, velocidade do vento e temperatura. Devido à sua natureza intermitente, a modelagem adequada dessas fontes de geração não convencionais é o primeiro passo a ser realizado em estudos para a implantação de unidades de GD renováveis, uma vez que o tratamento adequado das características estocásticas das fontes na modelagem leva a modelos com maior confiabilidade.

Diferentes técnicas probabilísticas foram usadas nas últimas décadas para modelar essas fontes intermitentes. Os modelos comumente utilizados incluem modelos de média móvel integrada, modelos auto regressivos e distribuições de probabilidade que compõem um processo aleatório indexado no tempo. Estudos como os realizados por Cargnelutti *et al.* (2004), Atwa *et al.* (2010), Cordeiro (2016) e Souza *et al.* (2018) sugerem o uso de técnicas probabilísticas como alternativa para representar incertezas na modelagem de fontes renováveis de energia, e devido a estas técnicas terem representações mais precisas das variáveis aleatórias.

Este artigo apresenta um estudo de avaliação probabilística de recursos de energia renováveis para verificar o ajuste dos dados de irradiância solar e velocidade do vento de quatro estações meteorológicas das cidades da Paraíba, o qual teve como objetivos: i) analisar minuciosamente as características dessas fontes intermitentes; ii) identificar o modelo mais adequado para descrever as séries de dados históricos de ambos os recursos investigadas; iii) prover parâmetros para modelagem das fontes intermitentes, baseados em base de dados reais advindos de quatro estações meteorológicas.

O conteúdo apresentado neste artigo está organizado da seguinte forma: na seção 2 são apresentadas e explicadas as etapas adotadas para gerar os modelos estatísticos das fontes renováveis, na seção 3 são descritos a avaliação da modelagem e os resultados dos recursos renováveis, e na seção 4 são feitas as conclusões.

## 2. MODELAGEM DE FONTES RENOVÁVEIS DE ENERGIA

A irradiância solar e a velocidade do vento são processos aleatórios que podem ser modelados por funções de densidade de probabilidade (FDP) ou função de massa de probabilidade (para processos discretos). Uma função de distribuição acumulada (FDA) é a probabilidade de uma variável aleatória ( $X$ ) ser menor ou igual a um valor específico de  $x$  e a FDP é uma função que pode ser usada para definir uma distribuição de probabilidade específica (Gentle, 2003).

Estudos relacionados a essa área explicam avaliações realizadas com séries de dados relacionadas a meses ou no máximo de alguns anos (Nunes, 2006), (Oliveira, 2016), mas para obter um modelo estocástico mais preciso, faz-se necessário selecionar dados históricos. Irradiância solar e velocidade do vento, que são objeto de estudo deste trabalho, são processos longos para serem representados em séries de dados contendo curtos períodos. Portanto, para gerar os modelos estatísticos que representam os processos estocásticos de irradiância solar e velocidade do vento, o primeiro passo é obter dados dos processos na forma de séries temporais. Assim, a partir das séries temporais, são realizados os ajustes às distribuições de probabilidade. O procedimento de ajuste é realizado a fim de se obter resultados generalizados que são mais confiáveis em simulações de processos estocásticos (Dolatabadi, Ebadi, Mohammadi-Ivatloo, 2019). Neste trabalho, Dez anos de dados de irradiância solar e velocidade do vento foram coletados em intervalos horários, nas cidades de Areia, Campina Grande, João Pessoa e Patos e utilizados para gerar os modelos estatísticos.

## 2.1 Processamento dos Dados Históricos

O procedimento para determinar quais funções de distribuição de probabilidade que podem representar os conjuntos de dados de uma medição estão resumidos no fluxograma da Fig. 1. Em seguida são descritos cada etapa do fluxograma para um melhor entendimento do procedimento.

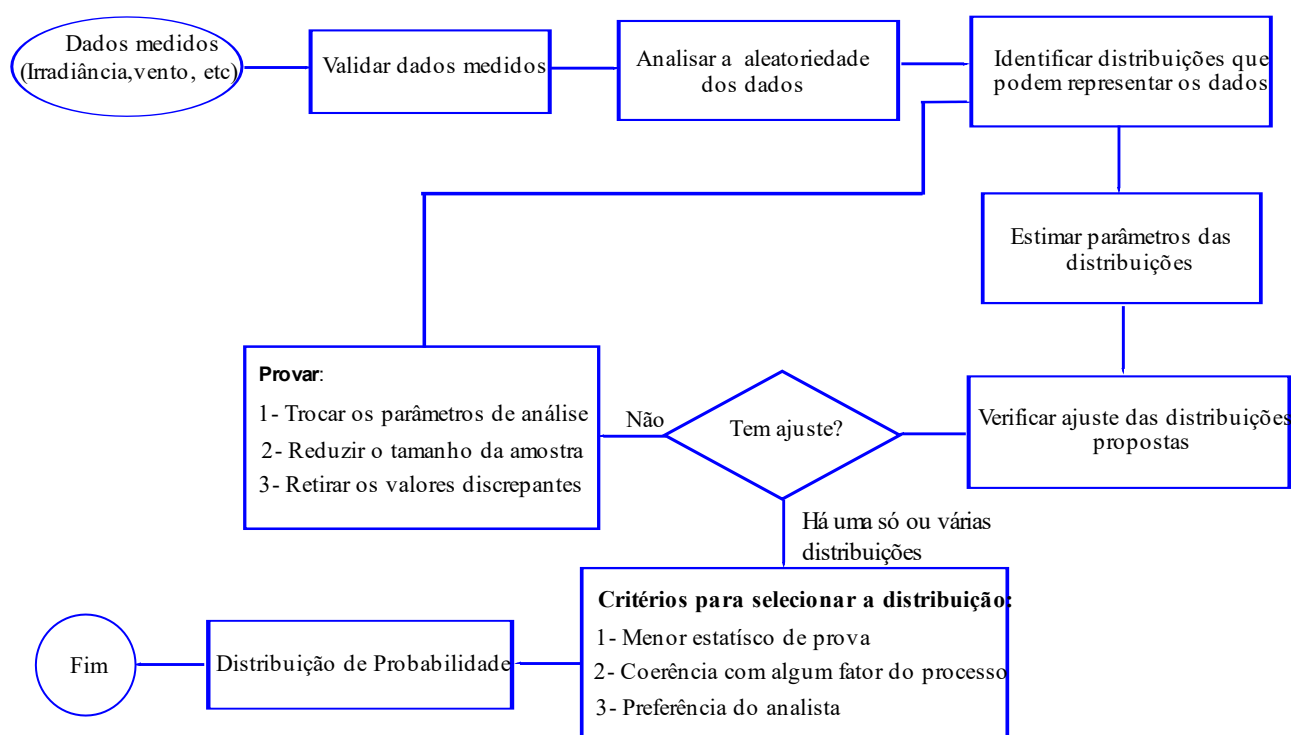


Figura 1: Procedimento para determinar qual FDP melhor representa os dados de uma medição.

Para o desenvolvimento dos modelos, dez anos de dados horários de irradiação solar em  $\text{kJ/m}^2$  e velocidade do vento em  $\text{m/s}$  para as quatro cidades foram cedidos pelo Instituto Nacional de Meteorologia (INMET) (INMET, 2018). Assim sendo, com o objetivo de desenvolver modelos de geração de insumos renováveis os dados foram tratados e optou-se por valores de irradiância solar ( $\text{W/m}^2$ ) devido a forma ser mais relevante nos estudos sugeridos (fluxo de carga, dimensionamento de sistemas, localização de fontes renováveis, entre outros). Os dados são de 2008 a 2017 e foram disponibilizados para este estudo mediante solicitação por e-mail. A seguir descreve-se detalhadamente as etapas mostradas na Fig. 1:

**Selecionar dados históricos que representam o fenômeno.** Para descrever um fenômeno aleatório na forma de um modelo estocástico, é necessário selecionar dados históricos que o representam. Ademais, é importante enfatizar que séries temporais desses processos com diferentes intervalos de amostragem podem ter FDP diferentes ou a mesma FDP com parâmetros diferentes. O intervalo de amostragem das séries temporais selecionadas pode ser definido pelas limitações do equipamento de aquisição de dados ou por uma decisão tomada em benefício do estudo que está sendo realizado. Para irradiância solar e a velocidade do vento, os dados medidos por hora são a forma mais comum apresentada na literatura especializada. A duração da série temporal também é importante porque, se for muito curta, o componente aleatório do processo poderá ser perdido.

Para as quatro cidades aqui consideradas, o período de estudo de dez anos é condensado em um ano típico em que a cada dia do ano é atribuído um conjunto de dados contendo os valores para esse dia, nos dez anos da série de dados original. E para cada dia a média dos valores foi calculada. A base de dados resultante contém valores que representam o período de 10 anos. A conversão dos dados de dez anos em uma média diária de um ano típico é realizada para reforçar

o conjunto de dados contra falhas resultantes de erros de leitura ou falta de dados no conjunto de dados original sem perder a característica estocástica do processo ao longo do tempo.

**Validar as informações.** Os conjuntos de dados resultantes da etapa 1 devem ser verificados quanto a inconsistências ou erros. Por exemplo, como o banco de dados de irradiação e velocidade do vento devem ter apenas valores positivos (característica natural dos processos), valores negativos devem ser excluídos. Da mesma forma, dados muito discrepantes, ou seja, muito diferentes ou estranhos com respeito aos outros dados que se tem em estudo são descartados. Existem diferentes critérios para descarte de dados a depender das características do conjunto, do processo estocástico em si e de possíveis erros humanos ou de falha de equipamentos. Por exemplo, para algumas aplicações, como fluxo de potência, valores de irradiação menores que  $10 \text{ W/m}^2$  podem ser considerados discrepantes visto que são praticamente irrelevantes a estudo a ser realizado. Esta etapa garante que o banco de dados seja composto apenas pelos dados coletados do fenômeno real em estudo, dados mal interpretados, erros de conversão e mau funcionamento do equipamento sejam eliminados.

**Analisar a aleatoriedade dos dados.** Modelos probabilísticos podem ser ajustados apenas ao processo aleatório. Por definição, um processo estocástico (aleatório) é um fenômeno estatístico que consiste em uma coleção de variáveis aleatórias ordenadas no tempo (Cox, 2017). É possível determinar a aleatoriedade de um conjunto de dados analisando a natureza do processo ou aplicando um teste de aleatoriedade. A irradiação solar e a velocidade do vento são processos aleatórios conhecidos, portanto, não há necessidade de aplicar um teste. O comportamento estocástico associado à natureza do problema é apresentado nos gráficos de boxplot da Fig. 2 para exemplificar.

Na Fig. 2 são apresentados os gráficos de caixas da irradiação solar média diária normalizada, agrupada por meses para os anos de 2012, 2013 e 2014 para ilustrar a diferença dos anos citados.

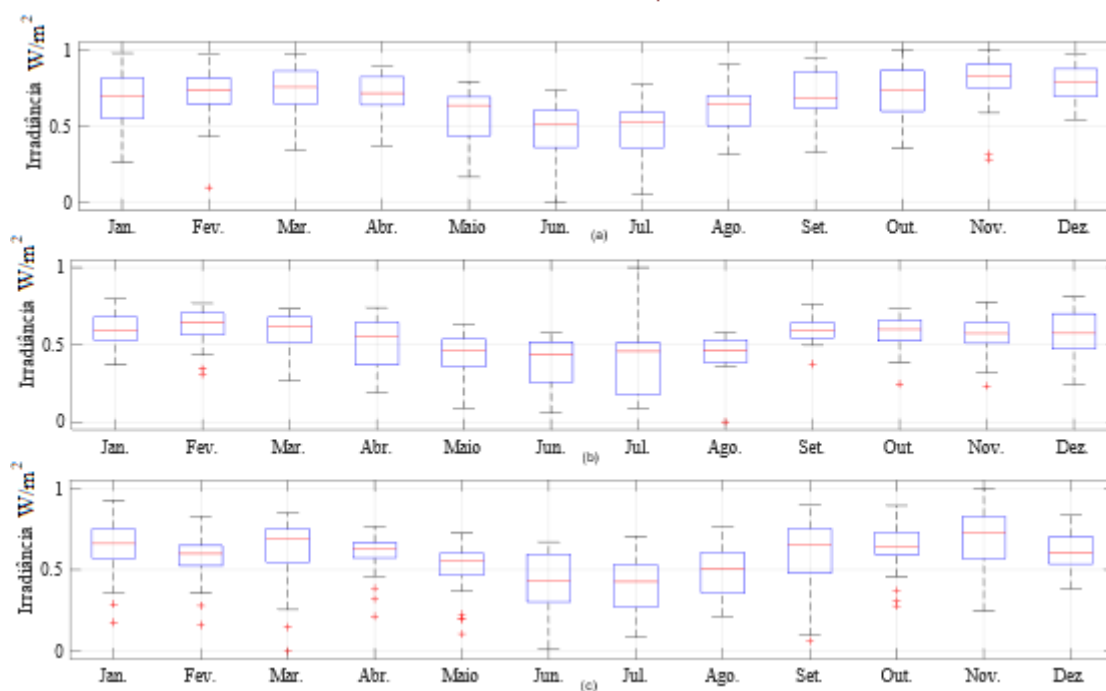


Figura 2: Comparativo entre os dados de irradiação solar (a) boxplot 2012, (b) boxplot 2013 e (c) boxplot 2014.

**Identificar as distribuições que podem representar o processo.** Nesta etapa são feitas algumas hipóteses de distribuições de probabilidade que podem representar melhor o conjunto de dados. Isso é feito considerando o tipo de variável aleatória (discreta ou contínua), a faixa de valores das variáveis, os tipos de dados (positivos ou negativos), a forma do histograma e o coeficiente de variação estatística (Pareja, 2009), (Zapata, 2005).

**Estimar os parâmetros das distribuições que podem representar o processo.** O cálculo dos parâmetros de algumas funções de distribuição pode apresentar dificuldades devido à complexidade matemática. Os parâmetros das distribuições podem ser estimados usando os seguintes métodos: métodos gráficos, método de momentos e método de máxima verossimilhança. De acordo com a literatura, o melhor método para calcular os parâmetros de distribuição é o método de máxima verossimilhança, porque os métodos gráficos são subjetivos e os métodos de momento não têm justificativa matemática (Pareja, 2009) (Cordeiro, 2016).

**Verificar o ajuste das distribuições propostas.** A visualização simples dos dados amostrais de uma variável em um histograma de frequência não é suficiente para concluir, dentre as várias funções conhecidas de distribuição de probabilidade, que melhor se ajustam aos dados dos estudos (Cargnelutti *et al.*, 2004) (Souza *et al.*, 2018). Portanto, é necessário usar testes de hipóteses, conhecidos como testes de aderência, para validar a adequação do modelo probabilístico a um conjunto de dados observado. Para verificar o ajuste das distribuições propostas, podem ser aplicados os testes de aderência Qui-quadrado, Kolmogorov-Smirnov (KS) e Anderson-Darling. Entre eles, o teste KS é o mais aplicado nos trabalhos encontrados na literatura (Cordeiro, 2016), (Souza *et al.*, 2018), (Pareja, 2009), (Zapata, 2005), (Law e Kelton, 2000) (Lima *et al.*, 2019).

**Determinar a distribuição que representa o processo.** Determinar quais funções de distribuição de probabilidade representam melhor o banco de dados em estudo: nenhuma FDP, múltiplas FDP ou uma FDP.

Se não houver uma função de distribuição que representa os dados, os seguintes procedimentos podem ser executados (Pareja, 2009), (Zapata, 2005):

- 1) Alterar os parâmetros de análise, alterando o nível de confiança;
- 2) Diminuir o tamanho da amostra se ela for muito grande;
- 3) Retirar os valores discrepantes e repetidos.

Se houver várias funções de distribuição de probabilidade, os seguintes critérios para escolher a distribuição podem ser adotados:

- 1) Escolher a distribuição de probabilidade que possui o maior valor-p de teste;
- 2) Escolher a distribuição de probabilidade que melhor se ajusta às condições naturais do problema em estudo;
- 3) Escolher a distribuição de probabilidade que melhor se ajusta ao projetista.

## 2.2 Método de Máxima Verossimilhança

O método de máxima verossimilhança é um dos procedimentos mais utilizados para obter estimadores para o tratamento do problema de estimativa com base nos resultados obtidos pela amostra. Deve-se determinar qual distribuição, dentre todas aquelas definidas pelos possíveis valores de seus parâmetros, com maior possibilidade de gerar uma amostra desse tipo. Assim, se, por exemplo, a distribuição da velocidade do vento é a distribuição Weibull, para cada combinação diferente de  $c$  e  $k$  existem diferentes distribuições Weibull. O estimador de probabilidade máxima escolhe o par de  $c$  e  $k$  que melhor explica a amostra observada. A função de probabilidade para um parâmetro genérico é dada por:

$$L(\theta) = \prod_{i=1}^n f(t_i, \theta) \quad (1)$$

em que:  $\theta$ : representa um único parâmetro ou um conjunto de parâmetros,  $L(\theta)$  é a função de verossimilhança e  $t_i$  é a amostra.

## 2.3 Teste de Aderência

O teste de aderência de Kolmogorov-Smirnov (KS) é um teste não paramétrico que tem como objetivo verificar se uma série de variáveis aleatórias (ou processo estocástico) tem determinada distribuição teórica, ou seja, averiguar se uma amostra pode ser considerada proveniente de uma população com determinada distribuição de probabilidade. Esse teste é baseado na distância máxima entre as probabilidades empíricas e teóricas obtidas para a função de distribuição de probabilidades em teste (Campos, 1983), (Razali e Wah, 2011), (Lima *et al.*, 2018):

A estatística de Kolmogorov-Smirnov para uma dada função distribuição acumulada  $F_0(x)$  :

$$D = \max |F_0(x) - S_n(x)| \quad (2)$$

em que:  $F_0(x)$ : é a distribuição hipotética,  $S_n(x)$ : uma distribuição observada em uma amostra de  $n$  observações (distribuição empírica).

Basicamente, este é um teste da hipótese nula. Assim, menores valores da estatística fornecem maiores níveis de significância (valor-p) e, conseqüentemente, maior evidência de não rejeição da hipótese nula, ou  $H_0$  e há, maior aderência dos dados à distribuição em teste. Portanto, quando o valor-p for maior que 5% a hipótese nula é aceitável, e os dados seguem a distribuição em que se está cogitando. Caso o valor-p seja menor que 5%, a hipótese é rejeitada, ou seja, os dados não seguem a distribuição cogitada (Campos, 1983), (Ferreira e Patino, 2015).

## 2.4 Distribuição Generalizada de Valores Extremos

A GEV combina em uma forma três tipos de distribuição de valores extremos: Gumbel, Fréchet e Weibull, também conhecidas como distribuições de valores extremos do tipo I, II e III. Esta distribuição tem sido usada para modelar uma grande variedade de extremos naturais, incluindo inundações, chuva, velocidade do vento, ondas e outras elevações, conforme Coles (2001). Matematicamente, a distribuição GEV é interessante porque seu inverso tem uma forma fechada e os parâmetros são facilmente estimados por momentos (Hosking, *et.al.*, 1990). As famílias I, II e III têm dois parâmetros: um da posição ( $\mu$ ) e outro da escala ( $\sigma$ ). As famílias II e III têm um fator de forma ( $\xi$ ). Cada uma das famílias pode ser combinada em uma única família de modelo com a FDP e a FDA, respectivamente 3 e 4:

$$g(z) = \frac{1}{\sigma} \left[ 1 + \xi \left( \frac{z - \mu}{\sigma} \right) \right] \left( \frac{-1}{\xi} \right)^{-1} \exp \left\{ - \left[ 1 + \xi \left( \frac{z - \mu}{\sigma} \right) \right] \left( \frac{1}{\xi} \right) \right\} \quad (3)$$

$$G(z) = \exp \left\{ - \left[ 1 + \xi \left( \frac{z - \mu}{\sigma} \right) \right] \left( \frac{1}{\xi} \right) \right\} \quad (4)$$

Graficamente, a distribuição de Fréchet corresponde a um modelo com uma cauda inferior finita e uma cauda superior infinita. Para a distribuição Weibull, a cauda superior é finita. Na Fig. 3 é apresentado o gráfico da função densidade de probabilidade para  $\xi = -0,3$  (Weibull),  $\xi = 0$  (Gumbel) e  $\xi = 0,3$  (Fréchet), com  $\mu = 10$  e  $\sigma = 2,5$ .

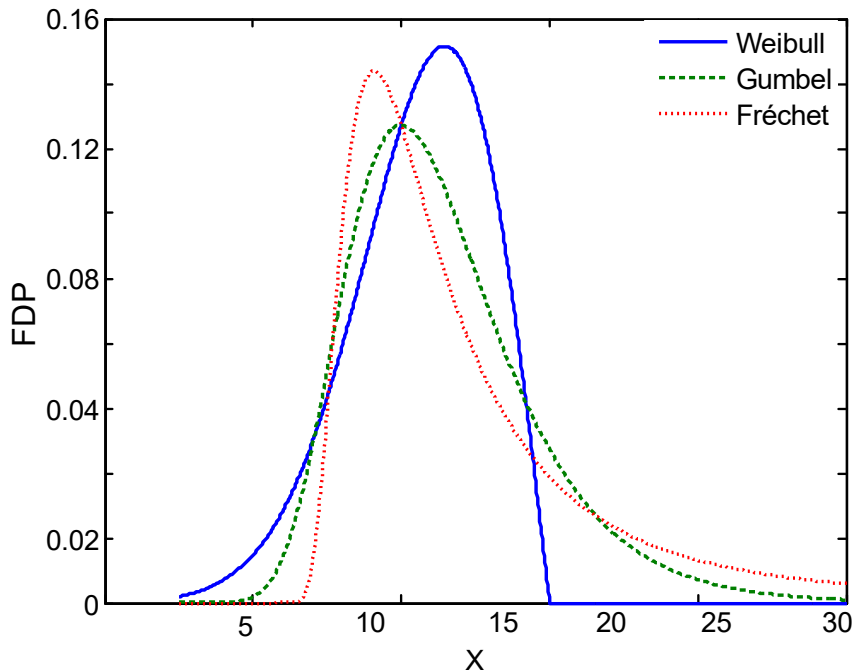


Figura 3: Funções distribuição de probabilidade GEV com diferentes parâmetros.

## 2.5 Distribuição Weibull

A distribuição Weibull é considerada pela literatura como uma das funções de densidade de probabilidade contínua mais estabelecidas. Nomeada após Waloddi Weibull, que em 1951 lançou um artigo descrevendo a distribuição em detalhes e propondo várias aplicações (Hosking, 1990). O campo de aplicação da distribuição Weibull é vasto e abrange estudos de vida útil do equipamento e estimativa de falhas. No entanto, como o vento é uma variável aleatória, com oscilação considerável no tempo, é a distribuição mais utilizada nesta análise. Entre os estudos pioneiros com essa distribuição estão os de Justus *et al.* (1978) que foi de suma importância para literatura. O sucesso da distribuição Weibull é justificado não apenas por sua eficácia, mas também pelo fato de existirem recursos gráficos que facilitam sua interpretação e por serem capazes de fazer previsões de precisão razoáveis, mesmo quando a quantidade de dados disponíveis é baixa.

Matematicamente, a distribuição Weibull é expressa por:

$$g(z) = \frac{k}{c} \left(\frac{v}{c}\right)^{(k-1)} \exp \left[-\left(\frac{v}{c}\right)^k\right] \quad (5)$$

$$G(z) = 1 - \exp \left[-\left(\frac{v}{c}\right)^k\right] \quad (6)$$

em que:  $c$ : o fator de escala em unidades de velocidade do vento;  $k$ : o fator de forma adimensional;  $v$ : a variável aleatória de velocidade do vento.

Os parâmetros  $c$  e  $k$ , conforme discutido em Justus *et al.*, (1978), podem ser determinados a partir da transformação da Eq. (6) em forma linear. Na Fig. 4 é mostrada a distribuição Weibull configurando o parâmetro  $k = 2$  e variando o parâmetro  $c = 1,5$  e  $3$ .

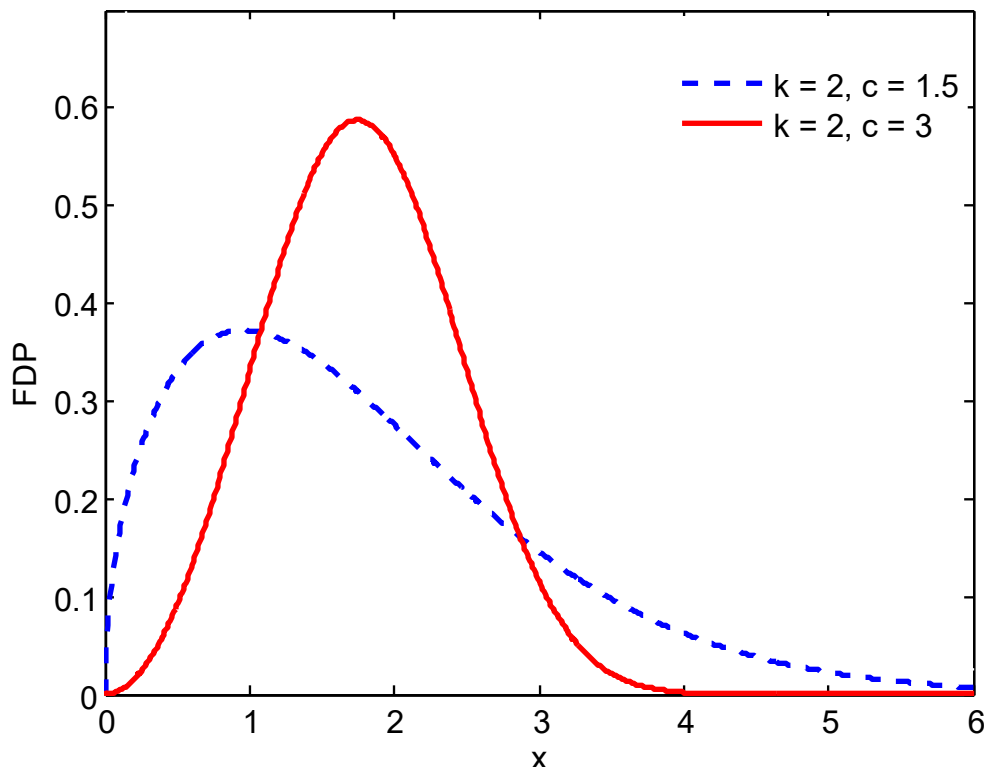


Figura 4: FDP da distribuição Weibull com diferentes parâmetros.

### 3. AVALIAÇÃO DA MODELAGEM DE RECURSOS RENOVÁVEIS

Um resumo dos resultados obtidos e as análises realizadas são relatados nesta seção tanto para irradiância solar quanto para a velocidade do vento. As etapas apresentadas na seção 2 foram seguidas para o banco de dados de irradiação solar e de velocidade do vento para as quatro cidades.

#### 3.1 FDP da Irradiância Solar

A irradiância solar é um processo que ocorre das 06:00 hs às 18:00 hs no Estado da Paraíba. Portanto, os valores irradiância solar abaixo de  $10 \text{ W/m}^2$  foram excluídos do banco de dados. O restante de cada conjunto de dados foi condensado em um ano típico. Na prática, o componente do ano foi eliminado do banco de dados e, para cada dia e hora, foi feita a média diária da irradiância solar. Portanto, os conjuntos de dados finais representam um ano típico de irradiância solar para as cidades. Em seguida, foram testadas várias distribuições, entre elas: Normal, Birnbaumsaunders, Exponencial, Valores Extremos, Gama, Valores Extremos Generalizado (GEV), Gaussiana Inversa, Logística, Lognormal, Nakagami, Poisson, *T Location Scale*, Rayleigh, Rician e Weibull. Os parâmetros de cada distribuição testada foram estimados pelo método da máxima verossimilhança. Posteriormente, os resultados foram submetidos ao teste KS e a hipótese nula não foi satisfeita para as distribuições: Normal, Exponencial, Valor Extremo, Gaussiana Inversa, Logística, Nakagami, Poisson, Rician e Rayleigh apresentaram vários erros para os quatro conjuntos de dados e, portanto, foram descartadas. As distribuições, Gama, GEV, *T Location Scale*, Lognormal e Weibull passaram no teste de aderência KS e, entre elas, a distribuição GEV obteve o melhor valor-p para os quatro conjuntos de dados referente a irradiância solar, ou seja, o valor-p foi maior que 5% e seus parâmetros foram ilustrados na Tab.1.

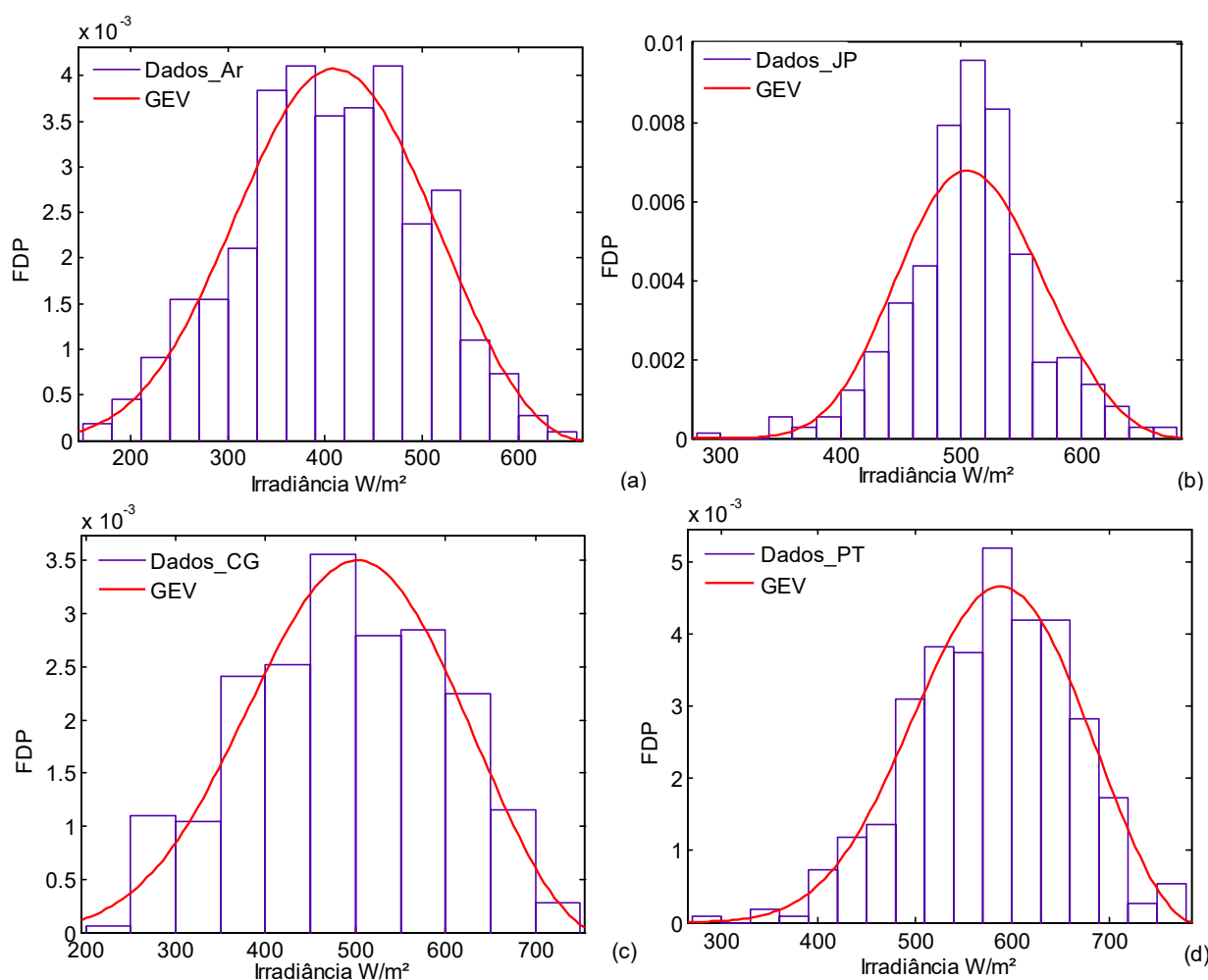


Figura 5: Distribuição GEV para irradiância nas quatro cidades: (a) Areia (b) João Pessoa (c) Campina Grande (d) Patos.

Tabela 1: Parâmetros da distribuição GEV e os resultados do teste KS.

GEV	K	$\sigma$	$\mu$	Média	valor-p
Areia	-0,32	95,80	374,13	405,41	0,92
Campina Grande	-0,35	113,00	455,82	490,49	0,83
João Pessoa	-0,27	56,48	487,96	508,27	0,16
Patos	-0,36	85,06	553,26	579,03	0,99

Certamente, o fato da distribuição GEV conter três parâmetros fornece mais um grau de liberdade em comparação com as outras distribuições testadas. Esse recurso explica a flexibilidade da GEV, bem como sua maior aderência aos dados de irradiância. No entanto, outra distribuição que também contém três parâmetros foi testada (*T Location Scale*) e não apresentou resultados semelhantes em comparação a GEV. Além disso, a Paraíba é conhecida por ter um longo verão e um inverno curto e é provavelmente por isso que a GEV representou tão bem os conjuntos de dados.

### 3.2 FDP da Velocidade do Vento

Para a velocidade do vento, os passos foram seguidos, conforme mostrados no fluxograma da Fig. 1. Como a velocidade do vento é um processo que ocorre 24 horas por dia, somente valores extremos (valores altos negativos e incomuns) foram excluídos dos conjuntos de dados iniciais e os valores restantes foram condensados em anos típicos.

Entre todas as distribuições, a Weibull apresentou uma boa alternativa para modelar os conjuntos de dados para as quatro cidades. O ajuste da distribuição Weibull aos conjuntos de dados é apresentado na Fig. 6 e na Tab. 2 é mostrado os parâmetros da distribuição e os resultados do teste de aderência KS.



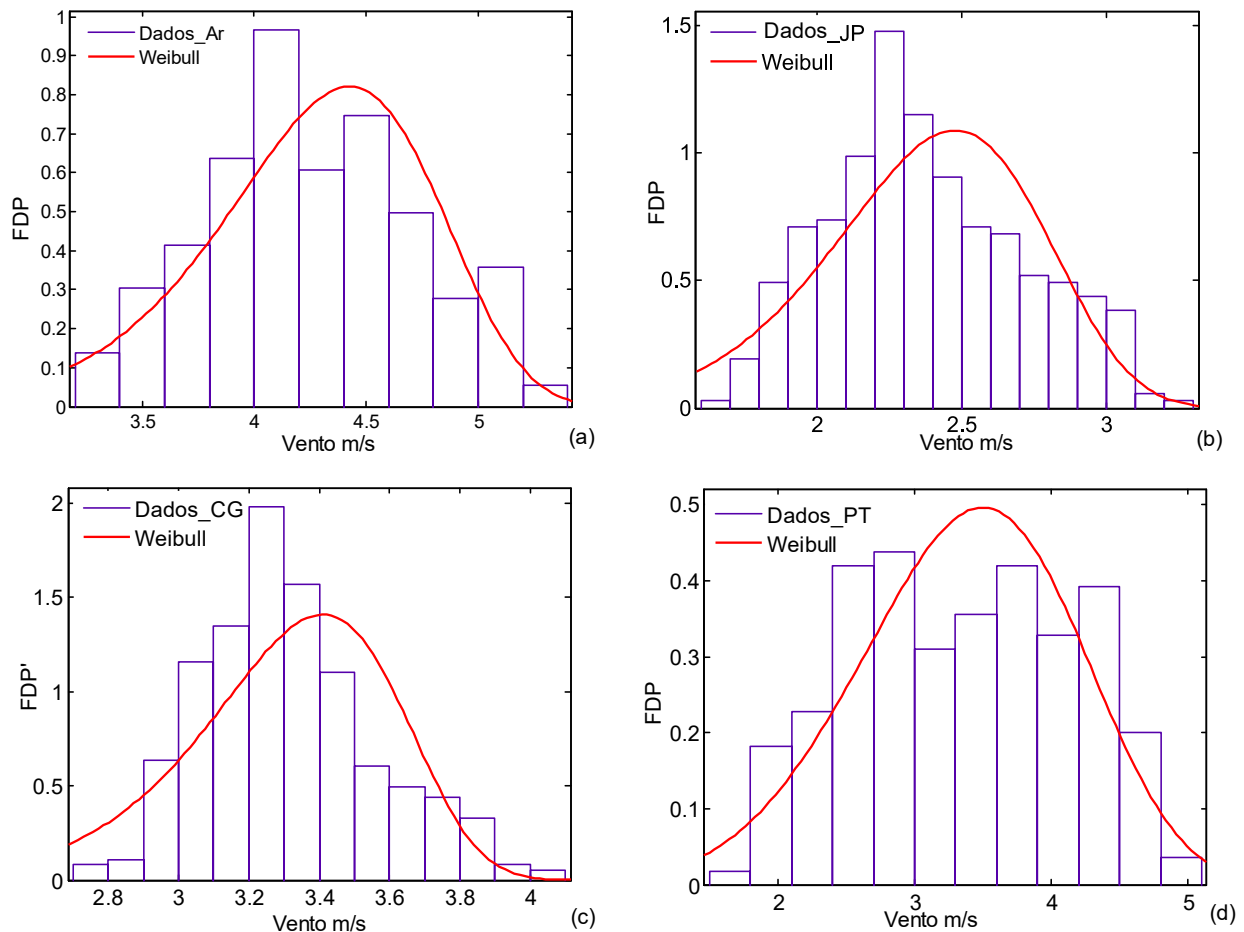


Figura 6: Distribuição Weibull para o vento nas quatro cidades: (a) Areia (b) João Pessoa (c) Campina Grande (d) Patos.

Tabela 2: Parâmetros da distribuição Weibull e os resultados do teste KS.

Weibull	$c$	$k$	Média	Valor-p
<b>Campina Grande</b>	3,43	13,10	3,29	0,08
<b>João Pessoa</b>	2,52	7,41	2,37	0,67
<b>Patos</b>	3,67	4,83	3,36	0,11
<b>Areia</b>	4,45	9,21	4,22	0,07

Para a velocidade do vento, a GEV não obteve bons resultados nos testes de aderência, tendo melhor adaptação à distribuição de Weibull.

#### 4 CONCLUSÃO

O conhecimento da distribuição diária média de irradiância solar e velocidade do vento é de suma importância ao otimizar o uso dessas fontes de energia renováveis. Neste trabalho, a irradiância e a velocidade do vento dos municípios de Areia, Campina Grande, Patos e João Pessoa de dez anos de dados foram melhor representadas pela distribuição de probabilidade de Valores Extremos Generalizados e distribuição de probabilidade de Weibull, quando comparadas a outras distribuições comumente apresentadas na literatura, apresentando um maior valor-p.

Com base no apresentado neste trabalho, os parâmetros de cada distribuição estimada podem ser utilizados em trabalhos com diversas aplicações para que estes tenham características e resultados mais realistas. Como outra contribuição a metodologia pode ser replicável em qualquer lugar do mundo, modificando apenas os dados de medição. Assim, a metodologia pode ser implementada para direcionar estudos como de localização ótima, confiabilidade dos módulos fotovoltaicos, sistema de proteção da qualidade de energia, harmônicos, tratamento de ruído entre outros.

## Agradecimentos

O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Código de Financiamento 001. Agradecem ao Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), a Coordenação de Pós Graduação em Engenharia Elétrica da UFCG (COPELE) e ao Instituto Nacional de Meteorologia (INMET) por ter cedido gentilmente os dados de irradiância solar e velocidade do vento de quatro estações meteorológicas da Paraíba.

## REFERÊNCIAS

- Ablein: <[http://www.inmet.gov.br/portal/index.php?r=home/pagepage=rede\\_estacoes\\_auto\\_graf](http://www.inmet.gov.br/portal/index.php?r=home/pagepage=rede_estacoes_auto_graf)>.
- Atwa, Y. M., El-Saadany, E. F., Salama e Seethapath, R., 2010. Optimal Renewable Resources Mix for Distribution System Energy Loss Minimization. IEEE Transactions on Power Systems, Vol. 25, n.1, pp. 360- 370.
- Campos, H., 1983. Estatística Experimental Não-Paramétrica. 4.ed. Piracicaba: Departamento de Matemática e Estatística – ESALQ. 349p.
- Cargnelutti, A, Matzenauer, R., Trindade, J.K., 2004. Ajustes de funções de distribuição de probabilidade à radiação solar global no Estado do Rio.
- Coles SG., 2001. An introduction to statistical modeling of extreme values. London: Springer, 208 p.
- Cordeiro, I.A., 2016. Modelo Generalizado de Valores Extremos Aplicado para Irradiação Solar e Análise de Confiabilidade de Sistemas Fotovoltaicos Residenciais. Dissertação, Universidade Federal de Campina Grande.
- Cox, D. R., 2017. The theory of stochastic processes. Routledge.
- Ferreira, J. C.; Patino, C. M., 2015. O que realmente significa o valor-p?. Jornal Brasileiro de Pneumologia, 41(5), 485-485.
- Dolatabadi, A., Ebadi, R., Mohammadi-Ivatloo, B. 2019. A two-stage stochastic programming model for the optimal sizing of hybrid PV/diesel/battery in hybrid electric ship system. Journal of Operation and Automation in Power Engineering, 7(1), 16-26.
- Gentle, James E., 2003. Random Number Generation and Monte Carlo Methods.
- Hosking, J.R.M. Wallis, J.R.; Wood, E.F., 1990. Estimation of the generalized extreme-value distribution b method of probability-weighted moments. Technometrics, v.27, p.251-261.
- INMET, Instituto Nacional de Meteorologia. Acesso em: 7 nov. 2018. Avail-
- Justus, C. G., Hargraves, W. R., Mikhail, A., Graber, D., 1978. Methods for estimating wind speed frequency distributions. Journal of applied meteorology, 17(3), 350-353.
- Law, A. M.; Kelton, W. D., 2000. Simulation modeling e analysis. New York: McGraw-Hill. 745 p.
- Lima, É. M., Junqueira, C. M., Almeida, R., Veloso, A., 2018. A. Modelagem probabilística do ruído em sistemas de distribuição de energia elétrica. Campina Grande.
- L-Moments: Analysis and estimation of distributions using linear combinations of order statistics, J. of Royal Statistical Society, B, 52(2), 105- 124.
- Morettin, P.A.; Bussab, W.O., 2004. Estatística básica. 5.ed. São Paulo: Saraiva, 526p.
- Myung, I.J., 2001. Tutorial on maximum likelihood estimation. Department of Psychology, Ohio State University, 1885 Neil Avenue Mall, Columbus, OH 43210-1222, USA.
- Nunes, F. A. H., 2006. Estimativa da Evapotranspiração de Referência em Campina Grande-PB. Dissertação em Meteorologia. UFCG.
- Oliveira, S. A., 2016. Geração Distribuída Fotovoltaica: Avaliação dos Impactos Técnicos e um Breve Estudo sobre Viabilidade Econômica. Dissertação de Mestrado. UFCG.
- Pareja, G. L. A., 2009. Fluxo de Potência em Redes de Distribuição de Energia Elétrica Considerando Incertezas, Tese de Doutorado, UESP.
- Razali, N. M.; Wah, Y. B., 2011. Power comparisons of shapiro-wilk, kolmogorov-smirnov, lilliefors and anderson-darling tests. Journal of statistical modeling and analytics, 2(1), 21-33.
- Souza, J.S, Alencar, M. R, Souza, B. A, Neves, W. A., 2018. Modelagem de Insumo de Geração Solar Usando Distribuições de Probabilidade. XXII Congresso Brasileiro de Automática-CBA, UFPB, João Pessoa-PB.
- Zapata, C. J., 2005. Análisis Probabilístico Y Simulación Colombia: Universidade Tecnológica de Pereira, Pereira.

## STATISTICAL MODELING OF UNCONVENTIONAL POWER SOURCES

**Abstract.** *This work presents a methodology for adjusting probability distributions to inputs for the generation of renewable energy sources (wind and solar). Based on statistical analyzes of the probability distribution function and the Kolmogorov – Smirnov adherence test. The proposed methodology is tested and the results are presented for four cities in Paraíba (Brazil). The results showed that, among the probability distributions mentioned in the literature, the distribution of Generalized Extreme Values is very flexible and can model ten years of global average solar irradiance data with p-values greater than 0.2 in all the cities analyzed. For the wind speed, the Weibull distribution is the one that best represents the database used with p-values above 0.07. The main contribution of this work is to present that this*

*methodology can be developed to guide studies that require data from renewable energy sources in the studied regions of Paraíba or other locations. For example, to carry out generalized simulations on load flows, dimensioning and localization of systems. In addition, it can be replicated for different databases and other variables.*

**Key words:** *Statistical Modeling, Generalized Extreme Values, Weibull.*